

# 自分の判断を説明できる自然言語処理システムを目指して

関根 聡<sup>1,a)</sup>

概要：自然言語処理技術は、機械学習やディープ・ラーニングの技術により格段の進歩を実現した。しかしながら、現状では、それらの技術はブラックボックスであり、なぜシステムがそのようなアウトプットを出したのか全く分からない。この問題は人工知能や自然言語処理技術が実世界に用いられる場面において非常に深刻な問題である。つまり、1) システムとしての説明責任、2) 100%ではないシステムのアウトプットの信頼性を人間が判断材料、3) 間違っていると人間が判断した時の再探索のための材料、としての重要な価値が説明によって提供される。このような説明をどうやって作るかということ自体が大きな研究テーマである。筆者は、3つの方向性を考えている。1) 1980年代に研究された推論システムを再現し、その頃の大問題であった知識ボトルネックの問題が Wikipedia や大規模コーパスで解決されたとして研究を進める方法、2) 推論システムと、ここ数十年で開発されてきた機械学習も含む統計的手法のハイブリットで実現する方法、3) ブラックボックスでアウトプットを生成した後に、同じような機械学習やディープ・ラーニングを利用して説明もブラックボックス的に作ってしまう方法。現在ではこの3つの方向性を考えてはいるが、どの方向性に行くにしても、言語知識、世界知識、推論知識が役に立つのは間違えないと考えており、そのような知識とそれを扱う手法についての研究を進めていきたいと考えている。

<sup>1</sup> 理化学研究所 革新知能総合研究センター／ニューヨーク大学  
RIKEN AIP, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 101-0027,  
Japan

a) [satoshi.sekine@riken.jp](mailto:satoshi.sekine@riken.jp)